Page 1, paragraph [0001], lines 2-6

"This application is a continuation of U.S. ~~application~~ Serial No. 09/620,658, filed July 20, 2000, and entitled "Network Performance ~~Mointoring~~ Monitoring in a Content Delivery Service". This application is related to U.S. Patent No. 6,108,703, issued August 22, 2000, filed May 19, 1999, which application was based on ~~provisional application~~ Serial No. 60/092,710, filed July 14, 1998."

Page 3, paragraph [0009], lines 19-22

"The ~~present invention provides a computer network comprising a large number of widely deployed Internet servers that form an organic, massively fault-tolerant infrastructure designed to serve Web content efficiently, effectively, and reliably to end users~~ disclosed subject matter is a method for Internet content delivery. The method begins by establishing a content delivery network at network locations, the content delivery network comprising a plurality of content servers for serving content resources. The plurality of content servers includes a plurality of subsets of content servers, each subject being located at one of a plurality of Internet data centers. For each Internet Protocol (IP) address block from which requests for content resources are expected to be received, the method generates a candidate list of Internet data centers to be used to service the requests for content resources. The candidate list of Internet data centers is generated using (i) geographic information from one or more Internet registry databases identifying a geographic location of the IP address block, (ii) BGP route information collected from BGP peers participating in BGP (BGP) sessions, (iii) autonomous system (AS) information, and (iv) data collected from one or more network performance metric tests. For the IP address block, the method selects at least one of the Internet data centers from the candidate list to be used to service the requests for content resources. The selected Internet data center for the IP address block is written into a network map. The selecting step is carried out concurrently for each IP address block from which requests for content resources are expected to be received such that the network map comprises the selected Internet data center for each IP address block. The network map is then provided to a domain name service (DNS) associated with the content delivery network. In response to a DNS query received at the domain name service associated

with the content delivery network, the network map is used to identify one of the Internet data centers from the candidate list to be used to service a request for a content resource.

Page 3, line 22, through page 4, line 2:

Please delete paragraph [0010] in its entirety.

Page 4, lines 3-5:

Please delete paragraph [0011] in its entirety.

Page 4, lines 6-9:

Please delete paragraph [0012] in its entirety.

Page 4, lines 10-12:

Please delete paragraph [0013] in its entirety.

Page 4, lines 13-15:

Please delete paragraph [0014] in its entirety.

Page 4, lines 16-21:

Please delete paragraph [0015] in its entirety.

Page 4, lines 22-23:

Please delete paragraph [0016] in its entirety.

Page 5, lines 1-8:

Please delete paragraph [0017] in its entirety.

Page 5, lines 9-20:

Please delete paragraph [0018] in its entirety.

Page 5, lines 21-23:

Please delete paragraph [0019] in its entirety.

Page 6, lines 1-10:

Please delete paragraph [0020] in its entirety.

Page 6, lines 11-15:

Please delete paragraph [0021] in its entirety.

Page 7, line 5:

"Figure 1 is a representative client-server system ~~in which the present invention is implemented~~ as known in the prior art;

Page 7, lines 8-9:

"Figure 3 is a high level diagram of a known global hosting system ~~according to the present invention~~;"

Page 7, lines 10-11:

"Figure 4 is a simplified flowchart illustrating a known method of processing a Web page to ~~modified~~ modify embedded object URLS ~~that is used in the present invention~~;"

Page 7, lines 12-13:

"Figure 5 is a simplified state diagram illustrating how the ~~present invention~~ the system of Figure 3 responds to ~~a~~ an HTTP request for a Web page;"

Page 9, line 9, through page 10, line 3 (paragraph [0028]):

Referring now to Figure 3, this operation is achieved by the hosting system ~~of the present invention~~ as described in U.S. Patent No. 6,108,703. As will be seen, the hosting system 35 comprises a set of widely-deployed servers (or server resources) that form a large, fault-tolerant infrastructure designed to serve Web content efficiently, effectively, and reliably to end users. The servers may be deployed globally, or across any desired geographic regions. As will be seen, the hosting system provides a distributed architecture for intelligently routing and replicating such content. To this end, the global hosting system 35 comprises three (3) basic types of servers (or server resources): hosting servers (sometimes called ghosts) 36, top-level DNS servers 38, and low-level DNS servers 40. Although not illustrated, there may be additional levels in the DNS hierarchy. Alternatively, there may be a single DNS level that combines the functionality of the top level and low-level servers. In this illustrative embodiment, the inventive framework 35 is deployed by an Internet Service Provider (ISP), although this is not a limitation of the present invention. The ISP or ISPs that deploy the inventive global hosting framework 35 preferably have a large number of machines that run both the ghost server component 36 and the low-level DNS component 40 on their networks. These machines are distributed throughout the network; preferably, they are concentrated around network exchange points 42 and network access points 44, although this is not a requirement. In addition, the ISP preferably has a small number of machines running the top-level DNS 38 that may also be distributed throughout the network."

Page 10, paragraph [0031], lines 15-22:

"According to the ~~present invention~~ disclosure in U.S. Patent No. 6,108,703, a given Web page (comprising a base HTML document and a set of embedded objects) is served in a distributed manner. Thus, preferably, the base HTML document is served from the Content Provider that normally hosts the page. The embedded objects, or some subset thereof, are preferentially served from the hosting servers 36 and, specifically, given hosting servers 36 that

are near the client machine that in the first instance initiated the request for the Web page. In addition, preferably loads across the hosting servers are balanced to ensure that a given embedded object may be efficiently served from a given hosting server near the client when such client requires that object to complete the page."

Page 11, line 19, through page 12, line 11 (paragraph [0034]):

"The ~~present invention~~ technique described in U.S. Patent No. 6,108,703 is not limited to generating the virtual server hostname by applying a hash function as described above. As an alternative and preferred embodiment, a virtual server hostname is generated as follows. Consider the representative hostname a1234.g.akamaitech.net. The 1234 value, sometimes referred to as a serial number, preferably includes information about the object such as its size (big or small), its anticipated popularity, the date on which the object was created, the identity of the Web site, the type of object (e.g., movie or static picture), and perhaps some random bits generated by a given random function. Of course, it is not required that any given serial number encode all of such information or even a significant number of such components. Indeed, in the simplest case, the serial number may be a simple integer. In any event, the information is encoded into a serial number in any convenient manner. Thus, for example, a first bit is used to denote size, a second bit is used to denote popularity, a set of additional bits is used to denote the date, and so forth. As noted above in the hashing example, the serial number is also used for load balancing and for directing certain types of traffic to certain types of servers. Typically, most URLs on the same page have the same serial number to minimize the number of distinguished name (DN) accesses needed per page. This requirement is less important for larger objects."

Page 12, paragraph [0035], lines 12-15:

"Thus, according to ~~the present invention~~ U.S. Patent No. 6,108,703, a virtual server hostname is prepended into the URL for a given embedded object, and this hostname includes a value (or serial number) that is generated by applying a given function to the URL or object. That function may be a hash function, an encoding function, or the like."

Page 19, line 14, through page 20, line 3 (paragraph [0060]:

"According to the ~~invention~~ disclosure in U.S. Patent No. 6,108,703, the virtual ghost names may be hashed into real ghost addresses using a table lookup, where the table is continually updated based on network conditions and traffic in such a way to insure load balancing and fault tolerance. Preferably, a table of resolutions is created for each serial number. For example, serial number 1 resolves to ghost 2 and 5, serial number 2 resolves to ghost 3, serial number 3 resolves to ghosts 2, 3, 4, and so forth. The goal is to define the resolutions so that no ghost exceeds its capacity and that the total number of all ghosts in all resolutions is minimized. This is done to assure that the system can take maximal advantage of the available memory at each region. This is a major advantage over existing load balancing schemes that tend to cache everything everywhere or that only cache certain objects in certain locations no matter what the loads are. In general, it is desirable to make assignments so that resolutions tend to stay consistent over time provided that the loads do not change too much in a short period of time. This mechanism preferably also takes into account how close the ghost is to the user, and how heavily loaded the ghost is at the moment."

Page 23, paragraph [0068], lines 4-9:

"The ~~inventive~~ global hosting scheme of U.S. Patent No. 6,108,703 is a way for global ISPs or conglomerates of regional ISPs to leverage their network infrastructure to generate hosting revenue, and to save on network bandwidth. An ISP offering the inventive global hosting scheme can give content providers the ability to distribute content to their users from the closest point on the ISPs network, thus ensuring fast and reliable access. Guaranteed web site performance is critical for any web-based business, and global hosting allows for the creation of a service that satisfies this need."

Page 23, paragraph [0069], lines 10-16:

"Global hosting according to ~~the present invention~~ U.S. Patent No. 6,108,703 also allows an ISP to control how and where content traverses its network. Global hosting servers can be set up at the edges of the ISP's network (at the many network exchange and access points, for example). This enables the ISP to serve content for sites that it hosts directly into the network

exchange points and access points. Expensive backbone links no longer have to carry redundant traffic from the content provider's site to the network exchange and access points. Instead, the content is served directly out of the ISP's network, freeing valuable network resources for other traffic."

Page 24, paragraph [0072], lines 4-10:

"Once inexpensive global hosting servers are installed at the periphery of an ISP's network (i.e., at the many network exchange and access points), content is served directly into network exchange and access points. As a result of this efficient distribution of content directly from an ISP's network, the ~~present invention~~ system as described in U.S. Patent No. 6,108,703 substantially improves Web site performance. In contrast to current content distribution systems, the ~~inventive~~ global hosting solution as described in U.S. Patent No. 6,108,703 does not require expensive backbone links to carry redundant traffic from the Content Provider's Web site to the network exchange and access points."

Page 35, paragraph [00105], lines 11-17:

"The problems solved are quite large, typically 150,000 CIDR blocks and 30+ data centers. The ~~roudned~~ rounded solutions are usually very close in both the total cost and the flow imposed on any data center to the cost and flow of the minimum-cost multi-path solution. More sophisticated deterministic algorithms could be applied here. In particular, there is an algorithm that guarantees that no data center receives more flow than that imposed by the optimal solution plus the flow out of one CIDR block. For a bipartite graph, the algorithm is a degenerative case for converting multi-path flows to single-path flows.